OIPE

RAW SEQUENCE LISTING   DATE: 11/08/2001
PATENT APPLICATION: US/09/982,704   TIME: 13:15:48

Input Set : A:\00032829.app
Output Set: N:\CRF3\11082001\I982704.raw

```
    3 <110> APPLICANT: KIY, THOMAS
    4         SCHULTZ, JOACHIM
    6 <120> TITLE OF INVENTION: CATHEPSIN-L, ITS PREPRO FORM AND THE CORRESPONDING
    7         PROPEPTIDE FROM CILIATES
    9 <130> FILE REFERENCE: 514489-3898
C--> 11 <140> CURRENT APPLICATION NUMBER: US/09/982,704
C--> 12 <141> CURRENT FILING DATE: 2001-10-18
   14 <150> PRIOR APPLICATION NUMBER: 08/981,957
   15 <151> PRIOR FILING DATE: 1998-04-13
   17 <150> PRIOR APPLICATION NUMBER: PCT/EP97/02388
   18 <151> PRIOR FILING DATE: 1997-05-09
   20 <150> PRIOR APPLICATION NUMBER: 19619366.4
   21 <151> PRIOR FILING DATE: 1996-05-14
   23 <160> NUMBER OF SEQ ID NOS: 16
   25 <170> SOFTWARE: PatentIn Ver. 2.1
   27 <210> SEQ ID NO: 1
   28 <211> LENGTH: 20
   29 <212> TYPE: PRT
   30 <213> ORGANISM: Paramecium tetraurelia
   32 <400> SEQUENCE: 1
   33 Gly Ala Glu Val Asp Trp Thr Asp Asn Lys Lys Val Lys Tyr Pro Ala
   34 1               5                   10                  15
   36 Val Lys Asn Gln
   37                 20
   40 <210> SEQ ID NO: 2
   41 <211> LENGTH: 10
   42 <212> TYPE: PRT
   43 <213> ORGANISM: Paramecium tetraurelia
   45 <220> FEATURE:
   46 <221> NAME/KEY: VARIANT
   47 <222> LOCATION: (1).(10)
   48 <223> OTHER INFORMATION: Xaa represents any amino acid
   50 <400> SEQUENCE: 2
W--> 51 Gly Ala Glu Val Asp Xaa Thr Xaa Asn Lys
   52 1               5                   10
   55 <210> SEQ ID NO: 3
   56 <211> LENGTH: 24
   57 <212> TYPE: PRT
   58 <213> ORGANISM: Paramecium tetraurelia
   60 <400> SEQUENCE: 3
   61 Asp Ser Ala Phe Glu Tyr Val Ala Asp Asn Gly Leu Ala Glu Ala Lys
   62 1               5                   10                  15
   64 Asp Tyr Pro Tyr Tyr Ala Ser Asp
   65                 20
   68 <210> SEQ ID NO: 4
   69 <211> LENGTH: 44
   70 <212> TYPE: DNA
```

ENTERED

```
 71 <213> ORGANISM: Artificial Sequence
 73 <220> FEATURE:
 74 <221> NAME/KEY: variation
 75 <222> LOCATION: (1)..(44)
 76 <223> OTHER INFORMATION: nucleotide 'w' can be either of the nucleotides
 77       'a' or 't'
 79 <220> FEATURE:
 80 <221> NAME/KEY: variation
 81 <222> LOCATION: (1)..(44)
 82 <223> OTHER INFORMATION: nucleotide 'h' can be either of the nucleotides
 83       'a' or 'c' or 't'
 85 <220> FEATURE:
 86 <221> NAME/KEY: variation
 87 <222> LOCATION: (1)..(44)
 88 <223> OTHER INFORMATION: nucleotide 'r' can be either of the nucleotides
 89       'a' or 'g'
 91 <220> FEATURE:
 92 <221> NAME/KEY: variation
 93 <222> LOCATION: (1)..(44)
 94 <223> OTHER INFORMATION: nucleotide 'y' can be either of the nucleotides
 95       'c' or 't'
 97 <220> FEATURE:
 98 <223> OTHER INFORMATION: Description of Artificial Sequence: primer 1
100 <400> SEQUENCE: 4
101 gcggggtacc ggwgchgaag thgaytggac wgataayaar aarg                    44
104 <210> SEQ ID NO: 5
105 <211> LENGTH: 12
106 <212> TYPE: PRT
107 <213> ORGANISM: Paramecium tetraurelia
109 <400> SEQUENCE: 5
110 Gly Ala Glu Val Asp Trp Asp Asn Lys Lys Val Lys
111   1               5                   10
114 <210> SEQ ID NO: 6
115 <211> LENGTH: 23
116 <212> TYPE: DNA
117 <213> ORGANISM: Artificial Sequence
119 <220> FEATURE:
120 <223> OTHER INFORMATION: Description of Artificial Sequence: primer 2
122 <220> FEATURE:
123 <221> NAME/KEY: variation
124 <222> LOCATION: (1)..(23)
125 <223> OTHER INFORMATION: nucleotide 'n' can be either of the nucleotides
126       'a', 'c','g' or 't'
128 <220> FEATURE:
129 <221> NAME/KEY: variation
130 <222> LOCATION: (1)..(23)
131 <223> OTHER INFORMATION: nucleotide 'r' can be either of the nucleotides
132       'a' or 'g'
134 <220> FEATURE:
```

```
       135 <221> NAME/KEY: variation
       136 <222> LOCATION: (1)..(23)
       137 <223> OTHER INFORMATION: nucleotide 'y' can be either of the nucleotides
       138      'c' or 't'
       140 <400> SEQUENCE: 6
W--> 141 tartanggrt artcyttngc ytc                                        23
       144 <210> SEQ ID NO: 7
       145 <211> LENGTH: 8
       146 <212> TYPE: PRT
       147 <213> ORGANISM: Paramecium tetraurelia
       149 <400> SEQUENCE: 7
       150 Glu Ala Lys Asp Tyr Pro Tyr Tyr
       151   1           5
       154 <210> SEQ ID NO: 8
       155 <211> LENGTH: 5
       156 <212> TYPE: PRT
       157 <213> ORGANISM: Paramecium tetraurelia
       159 <400> SEQUENCE: 8
       160 Gly Cys Asn Gly Gly
       161   1           5
       164 <210> SEQ ID NO: 9
       165 <211> LENGTH: 6
       166 <212> TYPE: PRT
       167 <213> ORGANISM: Paramecium tetraurelia
       169 <400> SEQUENCE: 9
       170 Cys Gly Ser Cys Trp Ala
       171   1           5
       174 <210> SEQ ID NO: 10
       175 <211> LENGTH: 31
       176 <212> TYPE: DNA
       177 <213> ORGANISM: Artificial Sequence
       179 <220> FEATURE:
       180 <223> OTHER INFORMATION: Description of Artificial Sequence: Primer(sense)
       182 <400> SEQUENCE: 10
       183 aggtcgtcat atgaatcttt atgcaaattg g                                31
       186 <210> SEQ ID NO: 11
       187 <211> LENGTH: 29
       188 <212> TYPE: DNA
       189 <213> ORGANISM: Artificial Sequence
       191 <220> FEATURE:
       192 <223> OTHER INFORMATION: Description of Artificial Sequence:
       193      Primer(antisense)
       195 <400> SEQUENCE: 11
       196 atcctcgagt cacttgtatt ggaagttag                                  29
       199 <210> SEQ ID NO: 12
       200 <211> LENGTH: 1276
       201 <212> TYPE: DNA
       202 <213> ORGANISM: Paramecium tetraurelia
       204 <400> SEQUENCE: 12
```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/982,704

DATE: 11/08/2001
TIME: 13:15:48

Input Set : A:\00032829.app
Output Set: N:\CRF3\11082001\I982704.raw

```
205 cattattagc agtcggttta atgatgttgt tgggagccag cctctacttg aacaacacat 60
206 aagaagtatc tgatgaaatc gatacagcaa atctttatgc aaattggaaa atgaaatata 120
207 acagaagata taccaactaa agagatgaaa tgtacagata caaggttttc acagacaacc 180
208 ttaactacat cagagctttc tatgaaagtc cagaagaagc cacattcact ttggaattga 240
209 atcaatttgc tgatatgagc taataagaat ttgcttaaac ctatttgagc ctcaaagttc 300
210 caagaacagc caaacttaat gccgccaatt ctaacttcta atacaagggt gcagaagtcg 360
211 attggactga caataagaag gttaagtatc cagctgttaa gaactaagga tcatgcggtt 420
212 catgctgggc cttctctgca gtcggagcac ttgaaatcaa cacagacatt gaactcaaca 480
213 gaaaatacga attatctgaa taagatttgg ttgactgctc aggaccatat gacaatgatg 540
214 gatgcaatgg tggatggatg gattctgctt ttgaatatgt tgctgacaac ggtttggctg 600
215 aagctaaaga ttatccatac actgctaaag atggaacctg caagacctca gttaaaagac 660
216 catacactca cgtctaagga ttcaaggata ttgactcatg cgatgaatta gcctaaacaa 720
217 tctaagaaag aacagtcgct gttgccgtcg atgccaatcc atggtaattc tacagaagtg 780
218 gtgtcctctc caaatgtact aaaaacttaa atcacggagt cgtccttgtt ggtgtttaag 840
219 ctgatggagc ttggaagatt agaaactcat ggggatctag ttggggagaa gctggtcaca 900
220 tcagacttgc cggaggtgat acttgcggta tctgtgctgc tccatctttc ccaattttag 960
221 gatgaagact ttgattattc atacatcaat ttacaacaat attagttatt tttaaactta 1020
222 agaaagactc ttgctgatgt tatcagtgaa ggattgaaaa aagtaggcac tctctaattg 1080
223 ggaggaggag ctgcatcaaa tgctccagct aaggcctaag ctccagctgc tgccaaataa 1140
224 gaggcaccaa agccagttga aaaggcccca gaaccagaag aagacgttga catgggtggt 1200
225 ttgtttgact gattatacat tttagtacat tcatatacat atattaaata ttttatcata 1260
226 aaaaaaaaaa aaaaaa                                               1276
```

229 <210> SEQ ID NO: 13
230 <211> LENGTH: 314
231 <212> TYPE: PRT
232 <213> ORGANISM: Paramecium tetraurelia
234 <220> FEATURE:
235 <221> NAME/KEY: PROPEP
236 <222> LOCATION: (1)..(109)
237 <223> OTHER INFORMATION: The position numbers for this sequence correspond
238   to -108 to 205 of Figure 2.
240 <400> SEQUENCE: 13

```
241 Met Met Leu Leu Gly Ala Ser Leu Tyr Leu Asn Asn Thr Gln Glu Val
242  1               5                  10                  15
244 Ser Asp Glu Ile Asp Thr Ala Asn Leu Tyr Ala Asn Trp Lys Met Lys
245             20                  25                  30
247 Tyr Asn Arg Arg Tyr Thr Asn Gln Arg Asp Glu Met Tyr Arg Tyr Lys
248         35                  40                  45
250 Val Phe Thr Asp Asn Leu Asn Tyr Ile Arg Ala Phe Tyr Glu Ser Pro
251     50                  55                  60
253 Glu Glu Ala Thr Phe Thr Leu Glu Leu Asn Gln Phe Ala Asp Met Ser
254 65                  70                  75                  80
256 Gln Gln Glu Phe Ala Gln Thr Tyr Leu Ser Leu Lys Val Pro Arg Thr
257                 85                  90                  95
259 Ala Lys Leu Asn Ala Ala Asn Ser Asn Phe Gln Tyr Lys Gly Ala Glu
260             100                 105                 110
262 Val Asp Trp Thr Asp Asn Lys Lys Val Lys Tyr Pro Ala Val Lys Asn
263             115                 120                 125
265 Gln Gly Ser Cys Gly Ser Cys Trp Ala Phe Ser Ala Val Gly Ala Leu
```

Input Set : A:\00032829.app
Output Set: N:\CRF3\11082001\I982704.raw

```
266       130                    135                    140
268 Glu Ile Asn Thr Asp Ile Glu Leu Asn Arg Lys Tyr Glu Leu Ser Glu
269 145                    150                    155                    160
271 Gln Asp Leu Val Asp Cys Ser Gly Pro Tyr Asp Asn Asp Gly Cys Asn
272                    165                    170                    175
274 Gly Gly Trp Met Asp Ser Ala Phe Glu Tyr Val Ala Asp Asn Gly Leu
275                180                    185                    190
277 Ala Glu Ala Lys Asp Tyr Pro Tyr Thr Ala Lys Asp Gly Thr Cys Lys
278            195                    200                    205
280 Thr Ser Val Lys Arg Pro Tyr Thr His Val Gln Gly Phe Lys Asp Ile
281        210                    215                    220
283 Asp Ser Cys Asp Glu Leu Ala Gln Thr Ile Gln Glu Arg Thr Val Ala
284 225                    230                    235                    240
286 Val Ala Val Asp Ala Asn Pro Trp Gln Phe Tyr Arg Ser Gly Val Leu
287                245                    250                    255
289 Ser Lys Cys Thr Lys Asn Leu Asn His Gly Val Val Leu Val Gly Val
290            260                    265                    270
292 Gln Ala Asp Gly Ala Trp Lys Ile Arg Asn Ser Trp Gly Ser Ser Trp
293        275                    280                    285
295 Gly Glu Ala Gly His Ile Arg Leu Ala Gly Gly Asp Thr Cys Gly Ile
296        290                    295                    300
298 Cys Ala Ala Pro Ser Phe Pro Ile Leu Gly
299 305                    310
```

302 <210> SEQ ID NO: 14
303 <211> LENGTH: 6
304 <212> TYPE: PRT
305 <213> ORGANISM: Paramecium tetraurelia
307 <400> SEQUENCE: 14

```
308 Glu Arg Phe Asn Ile Asn
309   1                5
```

312 <210> SEQ ID NO: 15
313 <211> LENGTH: 19
314 <212> TYPE: PRT
315 <213> ORGANISM: Paramecium tetraurelia
317 <220> FEATURE:
318 <221> NAME/KEY: VARIANT
319 <222> LOCATION: (1)..(19)
320 <223> OTHER INFORMATION: Xaa represents any amino acid
322 <400> SEQUENCE: 15

```
W-->  323 Glu Xaa Xaa Arg Xaa Xaa Val Phe Xaa Xaa Asn Xaa Xaa Xaa Ile Xaa
      324   1                5                    10                    15
W-->  326 Xaa Xaa Asn
```

330 <210> SEQ ID NO: 16
331 <211> LENGTH: 19
332 <212> TYPE: PRT
333 <213> ORGANISM: Paramecium tetraurelia
335 <220> FEATURE:
336 <221> NAME/KEY: VARIANT
337 <222> LOCATION: (1)..(19)

Use of n and / or Xaa has been detected in the Sequence Listing. Review the Sequence Listing to ensure a corresponding explanation is present in the <220> to <223> fields of each sequence using n or Xaa.

VERIFICATION SUMMARY
PATENT APPLICATION: **US/09/982,704**

DATE: 11/08/2001
TIME: 13:15:49

Input Set : **A:\00032829.app**
Output Set: **N:\CRF3\11082001\I982704.raw**

L:11 M:270 C: Current Application Number differs, Replaced Current Application Number
L:12 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:51 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:141 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6
L:323 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15
L:326 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15
L:341 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:16
L:344 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:16